



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Intraspecific host variation plays a key role in virus community assembly

Sallinen, Suvi ; Norberg, Anna ; Susi, Hanna ; Laine, Anna-Liisa

Abstract: Infection by multiple pathogens of the same host is ubiquitous in both natural and managed habitats. While intraspecific variation in disease resistance is known to affect pathogen occurrence, how differences among host genotypes affect the assembly of pathogen communities remains untested. In our experiment using cloned replicates of naive *Plantago lanceolata* plants as sentinels during a seasonal virus epidemic, we find non-random co-occurrence patterns of five focal viruses. Using joint species distribution modelling, we attribute the non-random virus occurrence patterns primarily to differences among host genotypes and local population context. Our results show that intraspecific variation among host genotypes may play a large, previously unquantified role in pathogen community structure.

DOI: <https://doi.org/10.1038/s41467-020-19273-z>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-200280>

Journal Article

Accepted Version

Originally published at:

Sallinen, Suvi; Norberg, Anna; Susi, Hanna; Laine, Anna-Liisa (2020). Intraspecific host variation plays a key role in virus community assembly. *Nature Communications*, 11:5610.

DOI: <https://doi.org/10.1038/s41467-020-19273-z>

Running title: Virus communities across host genotypes

INTRASPECIFIC HOST VARIATION PLAYS A KEY ROLE IN VIRUS COMMUNITY ASSEMBLY

Sallinen Suvi ^{*1}, Norberg Anna², Susi Hanna¹ & Laine Anna-Liisa²

¹ Organismal and Evolutionary Biology Research Programme, Viikinkaari 1 (PO box 65), FI-00014, University of Helsinki, Finland

² Department of Evolutionary Biology and Environmental Studies, University of Zürich, CH-8067 Zürich, Switzerland

Correspondence: suvi.sallinen@helsinki.fi, PO box 65, University of Helsinki, 00014-University of Helsinki, Finland, mobile: +358 50 311 5781, fax: +358 2941 57694

KEYWORDS: Coinfection, community assembly, genotype effects, host-parasite interaction, joint species distribution modelling, *Plantago lanceolata*, plant virus

ABSTRACT

Infection by multiple pathogens of the same host is ubiquitous in both natural and managed habitats. While intraspecific variation in disease resistance is known to affect pathogen occurrence, how differences among host genotypes affect the assembly of pathogen communities remains untested. In our experiment using cloned replicates of naïve *Plantago lanceolata* plants as sentinels during a seasonal virus epidemic, we find non-random co-occurrence patterns of five focal viruses. Using joint species distribution modelling, we attribute the non-random virus occurrence patterns primarily to differences among host genotypes and local population context. Our results show that intraspecific variation among host genotypes may play a large, previously unquantified role in pathogen community structure.

INTRODUCTION

Parasites constitute the majority of biological diversity on our planet ¹⁻⁴, and they influence both the demography and evolution of their host populations ⁵⁻⁷. Host susceptibility, pathogen infectivity, and environmental favourability have been identified as the corner stones of disease within the disease triangle framework ⁸. However, it is becoming increasingly clear that multiple infections within individuals are abundant ³, and have the potential to change the evolutionary and epidemiological trajectories of pathogens ⁹. Consequently, accounting for the diversity of infection is necessary to understand and predict disease dynamics and costs of infection for the host.

Understanding the determinants of the assembly and composition of pathogen communities is one of the key challenges in disease biology today. As a challenge it is analogous to the long-standing debate on the relative importance of biotic interactions versus external drivers of community dynamics. While some theories suggest species interactions to structure biological communities ^{10,11}, others highlight the importance of environmental drivers, including stress and disturbance on community dynamics ^{12,13}. To date, disentangling biotic processes from the abiotic ones has remained challenging ¹⁴. In recent years, pathogens are increasingly studied within a community ecological framework ¹⁵⁻¹⁹. Environmental variables and wider landscape context, such as human management, are linked to infection load, parasite diversity, and coinfection prevalence across multiple spatial scales ^{18,20-26}. The composition of parasite communities has also been linked to pathogen transmission mode, degree of host specialty, and life-cycle complexity ²⁷⁻²⁹, as well as host history, phylogeny, geographical range, longevity, and growth strategy ³⁰⁻³⁸. High parasite prevalence itself is a strong predictor of coinfections ^{9,39}. For

vector-borne diseases, positive co-occurrence is common for pathogens that share a vector or transmission site, or when vectors show preference for already infected individuals^{15,40,41}.

Co-occurrence of pathogens among host individuals is often non-random and coinfections can reach unexpectedly high levels^{15,20,23,42–44}. One of the key challenges is to determine how biotic interactions between hosts and their pathogens themselves shape these distributions. Under the community ecological framework, a host can be viewed as a resource patch and its resistance as a local filter that determines the pathogen community within that host¹⁸. Hosts are resistant against most pathogen species they encounter⁴⁵, and even for pathogens capable of infecting a host species, there is often considerable variation among individuals in their susceptibility^{7,46–50}. The effect of intraspecific variation in disease resistance on the dynamics of individual pathogens is well described^{51–54}. However, the importance of intraspecific host resistance variation for community assembly and diversity of species that exploit the host is only beginning to gain attention⁵⁵. Due to allocation costs associated with genetically-based resistance, a host resistant against a particular pathogen may be susceptible to others^{56,57}. On the other hand, limited evidence suggests that the same resistance loci may provide protection against several different pathogens⁵⁸. Pathogens attacking the same host may also compete for host resources (resource-mediated interaction), or interact via elicited host immune responses⁵⁹. Induced immunity by a first arriving pathogen may change the resistance phenotype, as immunosuppression of the host by the first arriving pathogen may facilitate establishment and replication of later arriving pathogens^{60–63}. On the other hand, cross-reactive immune responses elicited by the first parasite have the potential to suppress the success of later arriving parasites^{63,64}. These biotic interactions could result in non-random pathogen co-occurrence patterns across host genotypes. Variation in host resistance may be spatially structured with pronounced differences in resistance observed among host populations⁵³ and regions⁶⁵. Such spatially structured

resistance variation may also drive spatially structured co-occurrence patterns of pathogens exploiting the same host. Whether the host genotype is indeed a strong determinant of within-host parasite communities in the wild, and what the consequences of these within-host parasite community assembly processes are for host populations, remain unanswered^{18,66}.

Here, we study the importance of the host genotype in determining the structure of within-host virus communities. Viruses are in principle obligate parasites as they require a host for reproduction. A growing body of evidence has demonstrated that consequences of virus infection can shift along the pathogenic – mutualistic -continuum, even for the same interaction^{67,68}, and visually asymptomatic infections are common in wild plants³. Using cloned replicates of naïve *Plantago lanceolata* plants as sentinel traps placed in natural populations during a seasonal epidemic of viruses, we can tease apart the role of the host genotype from drivers that affect distribution of viruses within the local population context, which may include environmental variation, the local disease pool, host population structure and history, as well as local vector communities. Moreover, we aim to understand how biotic interactions among the viruses^{59–64} influence their community assembly.

We characterize the establishing virus communities using PCR detection⁶⁹. We first test whether the viruses occur in the same sentinel plant more often than would be expected based on their frequencies alone. In other words, we test whether virus co-occurrence patterns differ from expectations of a random distribution. We then employ a joint species distribution modelling (JSDM) framework⁷⁰, that allows us to tease apart the effect of local population context (consisting of unmeasured environmental variation as well as host population structure and history) on virus (co-)occurrences from host plant characteristics and host genotype. We can

account for the shared environmental responses of the target species, which makes the model a robust method also for sparse data ⁷¹. Using this approach, we are also able to capture signals of possible biotic community assembly processes from virus-to-virus association matrices after controlling for shared environmental responses of the viruses. The performance of JSDMs in relation to traditional, single-species distribution modelling (SDM) methods has recently been validated ⁷². The application of these kinds of multivariate statistical tools — typically used in community ecological analysis — to parasite data has the potential to reveal new insights of the determinants of parasite community assembly and composition ^{73,74}.

In this study, we ask: 1) Do we see more (or less, respectively) co-occurrences between the viruses than what would be expected solely based on their frequencies?; 2) Does the local population context affect the virus community composition?; 3) Do host genotypes differ in the virus communities they acquire, suggesting genotype-level variation in overall sensitivity to infection?; 4) *After* accounting for the aforementioned effects (2-3) of the local population context and plant host characteristics (including the host genotype), is there evidence of *residual* virus co-occurrence patterns across the entire data indicative of competitive or facilitative virus interactions?; and 5) Do these residual co-occurrence patterns vary among host genotypes indicating genotype-specific resistance responses affecting virus community structure? Our results indicate that while the population context also drives virus community assembly, host genotypes vary in the virus communities they acquire.

RESULTS

Detection of viruses in the field experiment

Out of the 320 sentinel host plants, 68 % were hosts to at least one virus over the study period. Three viruses were clearly more common in the sentinel plants: closterovirus in 120 individuals, betapartitivirus in 102 individuals, and capulavirus in 84 individuals; while caulimovirus and enamovirus were rare: in 10 and 5 individuals, respectively (Fig. 1a and Fig. 2). Out of the 217 infected individuals, 49 (23 %) hosted more than one virus, and in total we found 17 virus combinations, ranging from single infections to four of the five viruses found in the same plant (Fig. 2). Both overall virus prevalence and the composition of virus communities varied among plant genotypes and plant populations (Fig. 1a-b and Fig. 2).

Analysis of virus co-occurrence

We found significant non-random positive co-occurrences between species pairs capulavirus and caulimovirus as well as betapartitivirus and caulimovirus, when we analysed the complete data set (Fig. 3). When we analysed the co-occurrences separately for each host plant genotype, we found positive co-occurrences between betapartitivirus and caulimovirus on genotype 609_19, as well as between betapartitivirus and capulavirus on genotype 2818_6. We also found negative co-occurrences between betapartitivirus and closterovirus, as well as capulavirus and closterovirus on plant genotype 2818_6 (Fig. 3). When analysing co-occurrence patterns within each population, we found a significant positive association between capulavirus and closterovirus, and negative association between closterovirus and betapartitivirus in plant population 433. The expected and observed numbers of co-occurrence as well as the exact

probabilities for a greater or smaller number co-occurrences than expected for these species pairs are provided in Supplementary Table 3.

Joint species distribution models of virus communities

The model variants 2 and 3 performed almost equally well, as seen from their performance (Table 1). Model variant 1, excluding host plant genotype as a covariate, was clearly inferior. Model variant 2 also resulted in the smallest WAIC value, implying best predictive power. We did not detect any significant residual co-occurrence patterns between viruses after accounting for the effect of the local population context and host-related variables. We looked into this with sentinel plant level latent variables that are uniform across sentinel plant genotypes (model variant 2) as well as sentinel plant level latent variables that covary with sentinel plant genotype (model variant 3), and neither of these model variants captured virus co-occurrences with strong statistical support and their explanatory performances did not differ. Based on these results, we decided to consider the simpler model variant 2 as our best model.

The variance partitioning conducted for the model variant 2 revealed sentinel plant genotype to be the most important determinant for virus community composition (42% of variance explained, averaged over species; Fig. 4), followed by the local population context (29%; Fig. 4). The importance of variables differed between the viruses. Plant genotype explained most of the variation for capula- and caulimoviruses, while for enamovirus the sentinel plant genotype and the population context were almost equally important. For clostero- and betapartitiviruses the population context explained more variation in the data than plant genotype.

175

176 The importance of the random effect at the level of sentinel plant individuals differed between
177 the viruses, but followed roughly the same pattern: For capula-, caulimo- and enamovirus, the
178 sentinel plant individual random effect was minor, but for clostero and betapartitivirus, its effect
179 was slightly more pronounced (resulting in a total average effect of 14%). However, further
180 inspection revealed that none of the residual correlations between virus species gained strong
181 statistical support. Hence, we see no signal of potential biotic interactions between viruses after
182 taking into account the effects of fixed explanatory variables, i.e. the sentinel plant genotype,
183 size, signs of herbivory and local population context.

184

185 As expected, the predicted coinfections based on model variant 2 show similar patterns to what
186 we can see in the raw data (Fig. 2). When examining both the coinfection profiles (Fig. 2), and
187 the posterior mean estimates for the regression coefficients (Table 2), we see that capula- and
188 caulimovirus are much more likely to occur on sentinel plant genotype 609_19 (with posterior
189 mean estimate 2.47 for capula- and 0.67 for caulimovirus, Cap and Cau in Table 2, respectively,
190 that gained strong statistical support based on the 90% central credible interval). Other sentinel
191 plant genotypes were more dominated by single infections of closterovirus and betapartitivirus as
192 well as their co-occurrences. Thus, the overall structure of the virus communities among plant
193 genotypes was similar regarding the two most prevalent species closterovirus and
194 betapartitivirus, but sentinel plant genotype 609_19 hosted significantly more capulavirus, which
195 consequently also increases the probability of coinfections between capulavirus and other
196 viruses. Regarding caulimovirus, six out of the total ten of its occurrences were together with
197 capulavirus, and all of these co-occurrences were on sentinel plant genotype 609_19.
198 Closterovirus, betapartitivirus and capulavirus are tenfold more prevalent in our data in

comparison to caulimovirus and enamovirus, which can be seen in their dominance of the co-occurrence patterns in the community.

Sentinel plant size had a more minor effect on the community structure, as did signs of herbivory (Fig. 4), although both sentinel plant size and herbivory did have a minor positive effect with strong statistical support on the probability of occurrence of closterovirus (Table 2).

Our result for the same set of model variants fitted with less conservative priors for the latent part of the model show corresponding results to our main variants: model variant 1 is clearly inferior, whereas there is no big difference between variants 2 and 3. With model variants 2 and 3 we are able to detect one association with strong statistical support, between betapartitivirus and caulimovirus. For more details, see our Supplementary information on the joint species distribution modelling.

DISCUSSION

Understanding how pathogen communities are formed is a key challenge in understanding disease dynamics, as multiple infections can be significant drivers of epidemics as well as pathogen virulence and evolution^{9,18,19,75}. The host is expected to be a strong determinant in the formation of pathogen communities, as both theory and controlled experiments have demonstrated host resistance to be a key determinant of disease dynamics^{76–80}. Indeed, diversity of resistance in host populations could partly explain non-random co-occurrence patterns of

pathogens detected in wild plants^{15,20,23,44,46}. In our field experiment using sentinel plants of four genotypes, we found that most of the model-explained variation in virus occurrences was explained by the local population context and sentinel genotype (Fig. 4). Some viruses occurred significantly more or less together than would be expected based on their frequencies in both the full data set as well as when sentinel plant genotypes and local population context were analysed separately.

However, the results of our JSMD modeling (Table 1) indicate that the patterns evident in the co-occurrence analysis (Fig. 3) are influenced more clearly by the local population context and host genotype variation than by direct or indirect biotic interactions among the viruses. While disentangling host genotypic effects from other factors affecting pathogen communities has remained challenging, we were able to uncover the roles of these determinants of virus communities in wild hosts using naïve sentinel plants in wild plant populations.

Of the total amount of variation explained with our best model variant, the population context explained within-host virus communities to a large extent, although the proportion of explained variation varied among the viruses (Fig. 4). Drivers that could vary among our plant population include abiotic variables which we did not explicitly record as many more plant populations would be needed to tease apart relevant variation in local population context for virus communities. These drivers are often found to filter parasites according to their niche preferences from the regional disease pool into local populations, thereby playing a major role in how within-population and -host parasite communities are formed^{20,22,26}. In addition to abiotic variables, the local *P. lanceolata* populations are likely to differ in biotic factors including plant species community composition and abundance of suitable vectors which may be linked to virus

prevalence and diversity^{15,20}. The local population context further includes any differences in population dynamics and trajectories, such as historical pathogen pressure, which may vary among these populations⁸¹. Albeit non-significant, the effect of sentinel plant individual on the (co-)occurrences of the viruses can be attributed to some unmeasured abiotic or individual-related variables, which may influence the (co-)occurrences of the viruses.

While there are multiple studies investigating within-host parasite communities^{44,69,73,74,82}, to our knowledge the effect of host genotype on the assembly has rarely been tested experimentally in wild systems, or with multiple parasites simultaneously. In our data, sentinel plant genotype accounted for most of the variation in virus occurrences of the total variation explained in the JSDM model. Indeed, both virus occurrence, and the acquired virus communities varied among the four *P. lanceolata* genotypes. In particular, sentinel plant genotype 609_19 had greater infection prevalence and diversity of viruses than the other genotypes (Fig 2). As our model controlled for the effect of sentinel plant size and level of herbivore damage, such host genotype-level differences may reflect variation in constitutive resistance, such as resistance genes, among the plant genotypes. The natural *P. lanceolata* populations in the Åland Islands contain considerable phenotypic variation in resistance against powdery mildew *P. plantaginis*^{83,84}, and while resistance against viruses in this system is not well understood, an exceptionally diverse repertoire of candidate loci (Nucleotide-binding leucine-rich repeat; NLRs) that confer resistance against a broad range of pathogens, have been characterized in *P. lanceolata* (Laine, personal communication). Uncovering both phenotypic and molecular level virus resistance in this system is an important avenue of future research. Spatially structured variation in resistance is characteristic of natural host-parasite systems^{53,85–87}, and based on our findings, intraspecific variation in disease resistance in a host population may play a large, previously unquantified

role, in the non-random distribution of co-occurring pathogens that have been detected in previous studies^{15,20,23,44,46}.

Intraspecific variation in traits other than resistance could also generate the differences we observe. To confirm which traits are involved, future studies should explore in more detail the ecological outcomes of these interactions, and their molecular underpinning. It is highly plausible that the host genotype could indirectly affect virus occurrences via their attractiveness or resistance against vector herbivores^{88,89}. Vector preference for infected hosts^{41,90} could also influence virus co-occurrence patterns. Transmission mode is often found to be critical for how pathogen communities are formed^{15,40,91,92}, and reciprocally, the amount of genotypic variation within a host population may explain the abundance and composition of herbivory community present⁸⁹.

A community of pathogens could be shaped by both direct and indirect pathogen–pathogen associations: reaction triggered by an earlier arrival could either induce or suppress resistance against later arriving pathogens, or within-host competition could favor one pathogen over the other. Evidence for both negative and positive pathogen–pathogen interactions have been reported in studies of multiple infections^{19,59,62,63,93,94}. Although we find both positive and negative co-occurrence patterns among the viruses, these are largely explained by local population context and host genotype. After controlling for these in our model, we do not find strong statistical support for signals of associations among the viruses, as would be expected if arrival by one would decrease or increase the arrival probability of another. Hence, our results do not support the hypothesis that virus–virus interactions — either direct or those mediated by host immunity — would be the key drivers of virus community assembly at the within-host level in

294 this system. However, our sample size could be insufficient to detect such interactions as some
295 of our viruses are rare, and their arrival probability to the sentinel plants is also subject to
296 random processes. In addition, we only accounted for a subset of all possible pathogens infecting
297 plants in this system, thereby potentially missing some influential members of the community.
298 Furthermore, the effects of induced immunity triggered by a first arriving pathogen may be
299 short-lasting^{63,95} and therefore undetectable with the timescale of this experiment. Induced
300 immunity could play a more important role among viruses of the same genus or strains of the
301 same virus species, where the famous phenomenon of cross-protection is more often recorded
302^{63,96} and as is predicted by theory⁷⁵. Given that the variants with less conservative priors detected
303 a significant positive interaction between betapartitivirus and caulimovirus, we conclude that our
304 study design was successful in capturing the effects of the host genotype, but larger-scale
305 investigations would be required to detect signals of virus–virus interactions.

306
307 In our experimental design we kept the plants in their pots which meant these plants experienced
308 different rooting environments than the wild plants but allowed us to standardize some factors
309 (e.g., soil medium). However, this approach allowed us to control for this level of variation in
310 our data. Our approach may have affected vector preferences as visual presentation of the plants,
311 in addition to other cues, is important for vector dynamics⁹⁰. Nonetheless, transmission of all
312 five focal viruses to the sentinel plants did occur. Whether the virus prevalences we detected
313 with our approach are in line with infections of wild plants is difficult to assess, given that virus
314 prevalences vary greatly among populations in the Åland Islands (0-64 %) ⁶⁹. Overall, our study
315 does not only highlight the importance of the host genotype, but also the need for further
316 research on other aspects of virus ecology. Although we have placed the current work into a
317 context of pathogens, viruses may also have neutral or positive effects on the hosts despite their

parasitic lifestyle⁶⁷. While knowledge of virus diversity and roles of viruses in wild populations is increasing^{3,44,67}, research at the community scale remains scarce¹⁸.

Here, we have quantified the importance of intraspecific host plant variation on how within-host virus communities assemble by using sentinel plants in natural populations during a seasonal epidemic, which allows teasing this factor apart from other drivers of virus occurrence. Applying JSDMs to interpret the effects of host genotype and local population context, we find that while the population context has a strong influence on virus communities within individual hosts, not accounting for the host genotype might underestimate the role host genotypes have in generating variation in pathogen communities. Such variation in within-host pathogen diversity may have far reaching implications for all key aspects of disease: transmission, virulence suffered by the host, and pathogen evolution. With these results, we are one step closer to binding together the different spatial scales and processes that underpin pathogen metacommunities.

METHODS

Study species

Plantago lanceolata is a globally occurring perennial herbaceous plant⁹⁷. It is an obligate outcrosser with wind-dispersed pollen, also capable of vegetative reproduction⁹⁷. In the Åland Islands, SW of Finland, it typically grows on dry meadows, forming a network of approximately 4000 small connected populations⁸¹. The size and location of the populations have been

monitored since the early 1990s as a part of the metapopulation studies of the Glanville fritillary butterfly and powdery mildew *Podosphaera plantaginis*^{81,98}. In the Åland Islands, *P. lanceolata* also hosts a diverse community of viruses that vary in their occurrence among *P. lanceolata* populations and among the individuals within populations⁶⁹. We used five recently characterized viruses from the Åland Islands, to study within-host viral communities⁶⁹: *Plantago lanceolata* latent virus in genus Capulavirus and *Plantago lanceolata* caulimovirus in genus Caulimovirus with DNA-genomes, and *Plantago betapartitivirus* in genus Betapartitivirus, *Plantago enamovirus* in genus Enamovirus, and *Plantago closterovirus* in genus Closterovirus with RNA-genomes. The viruses are hereafter referred to by their genus for understandability. These viruses were initially identified from *P. lanceolata* in the Åland Islands by sequencing plant small RNAs⁶⁹. Plants use RNA-silencing mechanism and produce short interfering RNA (SiRNA) molecules in a defense response against viral infection⁹⁹. Hence, these viruses trigger an active defense response in *P. lanceolata*. Also, although not directly demonstrating their pathogenic nature, Susi *et al.*⁶⁹ found that plants with virotic symptoms (necrotic spots/yellow color) are more likely to carry a virus infection. Currently, the detailed transmission dynamics and vector species, as well as the viruses' distribution outside the Åland Islands remain unknown. More detailed information of the virus families is compiled in Supplementary Table 1.

Field experiment with sentinel plants of different genotypes

To study the effect of plant host genotype on the variation of within-host virus communities, we set up an experiment using sentinel trap plants in natural populations of *P. lanceolata* in the Åland Islands. To obtain genetically uniform plant material, we cloned four greenhouse-grown maternal *P. lanceolata* plants into 80 replicates each. The maternal plants originate from natural

P. lanceolata populations in the Åland Islands, and were grown from seeds in an insect free greenhouse at the University of Helsinki. The plants are expected to represent four different genotypes (ID:s 609_19, 4_13, 511_14, 2929_6), as their maternal plants originated from distant populations 7-40 kilometres apart. Their resistance against viruses is currently unknown, but they represent different mildew resistance phenotypes as has been confirmed during laboratory maintenance of *P. plantaginis*. The maternal plant individuals used in the experiment were confirmed to be free of target viruses, that would have been the result of seed borne infection, by PCR-testing using specific primers. Each maternal plant was cloned into 80 replicates by placing maternal plants on pots containing vermiculate and kept on a tray containing fertilized water.

After one month, the roots grown from the maternal plant's pot through to the vermiculate were cut. After another month, new plants shooting from the cut roots in the vermiculate were separated and individually planted into 10 cm × 10 cm pots containing an equal amount of sand and potting soil. After two additional months in the greenhouse, during the last week of May 2017, the plants were taken to the Åland Islands and placed into four *P. lanceolata* populations (ID:s 877, 9031, 433, 3302; Fig. 1c). The populations were selected for the study as they represent different parts of the Åland Islands, were remote to humans, and large enough to host a field-experiment. These populations were different from the ones the maternal plants used for cloning originated from. These four populations were included in the analyses as a categorical variable to capture 'local population context' (local temperature, vectors, plant communities etc.) that may influence virus distributions among *P. lanceolata* populations in the Åland Islands.

Twenty replicates of each sentinel plant genotype were placed into each of the four *P. lanceolata* populations resulting in 80 plants per population, and 320 plants altogether. The plants were kept

in their pots for the duration of the experiment, and they were placed in a random order among natural vegetation and reshuffled three times per week to avoid within-population spatial effects. The plants were kept separated from the local soil on plastic freezer boxes and watered when necessary. Signs of herbivory (holes, bite marks, and thrip damage) were recorded after two weeks of exposure, and again after seven weeks of exposure. Plant size was measured during the first week of exposure by counting the number of leaves, and by measuring length and width of the longest leaf. Based on these measurements we calculated plant size by using the equation $n \times A$, where n is the number of leaves, and leaf area A is calculated using the equation of ellipse area: $A = \pi ab$, where a is a half axis of the width of the longest leaf, and b is the half axis of the length of the longest leaf. For those 13 plants missing measurement data, an average over all recorded values for all plants was used, in order to not to lose any virus occurrence data from the analysis.

Nucleic acid extractions and virus detections with PCR

To detect the viruses infecting plants during the growing season, leaf samples were collected for nuclear acid extractions after two weeks and again after seven weeks of exposure to the natural virus and vector communities. Samples were collected from a single leaf of similar age (young but large enough for sampling) from each plant. For DNA extraction, we collected a 1 cm² piece of leaf from each plant. Samples were stored in -20 °C until DNA extraction with E.Z.N.A. Plant Kit (Omega Biotek, USA) at the Institute of Biotechnology at University of Helsinki. For RNA extractions, 3 cm² leaf samples were collected, immediately deep-frozen in liquid nitrogen, and stored in -80°C before RNA-extraction. Total RNA was extracted using phenol-chlorophorm extraction with a modified method from Chang *et al.* (1993). Two additional phenol cleaning

steps prior chloroform cleaning of the RNA were performed. In the additional cleaning steps, we used 800 ml of equal volumes of phenol solution (pH 4.5) and chloroform-isoamylalcohol, mixed with isolation buffer containing the sample, vortexed, and centrifuged for phase-separation in 14 800 rpm for 15 minutes. For the PCR detection of the RNA viruses, RNA was translated into cDNA. For reverse transcription, we used 2 ng of total RNA, mixed with 2µl random hexamer primers (Promega) and sterile nuclease free water in 17,125 µL volume incubated for 5 minutes in 70 °C. Subsequently, 1µL Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT; Promega Corporation, USA), 5 µL M-MLV RT buffer, 1.25 µL of dNTP (10mM) mix, and 0.625 µL of RiboLock RNaseinhibitor were added and the 37.41 µL reaction mix was incubated in 37°C for 60 minutes. For virus detection PCR, we used specific primers ^{69,101} as well as two additional primer pairs for capulavirus (PiLVi2_forward_1 5' GTGTTTAACAATGAAGTGAGCC 3' and PiLVi2_reverse_4 5' AATCCATCCACACATCCAATC 3') and caulimovirus (forward primer 5'AGGAGATGCCCATACTTTACC 3' and reverse primer 5' GACTTGCCAGAACCTGATTTAC 3'). PCR reactions to detect viruses were performed in final volume of 10 µL containing of 1-3 µL of DNA or cDNA, and GoTaq Green® polymerase 5x Mastermix (Promega Corporation, USA) according to manufacturer's instructions. Samples were subjected to initial denaturation in 95 °C for 2 min, following 35 cycles of denaturation in 95 °C for 40 s, annealing 53–60 °C for 40 s, and extension 72 °C for 1 min with a final extension step of 72 °C for 5 min. The full protocol with virus specific PCR conditions is described in the Supplement (section 'PCR-detection of viruses'). The amplicons were resolved on a 1.2-1.5% agarose gel and visualized using Gel Doc XR System (Bio-Rad Laboratories, Inc., USA).

Statistical analysis

437

438 For all the statistical analysis, we pooled the detected occurrences of the five focal viruses over
439 the two timepoints of sampling by collapsing the occurrence data so that each sentinel plant had
440 one observed virus community. Only when a sentinel plant had not been infected by a certain
441 virus in either of the timepoints accounted as an absence of the virus while infection in one or
442 both timepoints was accounted for as virus presence. To understand whether the co-occurrence
443 of viruses differs from expected co-occurrences calculated solely from the prevalences of these
444 viruses, we first analysed the co-occurrence patterns both in the full data set as well as separately
445 for each sentinel host genotype and plant population (Fig. 3). We used the R package ‘cooccur’
446 ¹⁰² and its identically named function, and applied a probabilistic model ¹⁰³ which calculates
447 expected frequencies of species co-occurrences based on a distribution of random, independent
448 species. By comparing the expected and observed co-occurrences the applied algorithm gives the
449 probabilities of co-occurrence greater than or less than what is observed in the data analytically,
450 without relying on randomisations or test statistics, under the condition that the probability of
451 occurrence for a species at each sentinel plant is equal to its observed frequency among all the
452 sentinel plants, i.e. in this case the prevalence of the virus¹⁰³.

453

454 For addressing our study questions about the effects of host genotype and characteristics as well
455 as local population context on the (co-)occurrence patterns of the viruses, as well as the possible
456 signals of biotic interactions between the viruses on virus community assembly, we applied a
457 joint species distribution modelling (JSDM) framework ‘Hierarchical Modelling of Species
458 Communities’ (HMSC¹⁰⁴), which is a multivariate Bayesian hierarchical generalised linear latent
459 variable model. Essentially, HMSC is a multivariate generalised linear model, enabling the
460 modelling of the whole community of viruses as opposed to fitting individual single-species

distribution models¹⁰⁵. In addition, HMSC is a latent variable model⁷⁰. Latent variable models include unobserved, i.e. latent predictors, which are typically included to model correlation, or to account for missing predictors⁷⁰. Hence, in this context, the latent variables are random effects that model the co-occurrences between species due to either biotic interactions or some other effects not included in the fixed part of the model, such as unmeasured effects of the environment. For a more detailed description of JSDMs and latent variable models, please see the comprehensive review by Warton *et al.*⁷⁰.

The structure of the HMSC modelling framework is described in detail by Ovaskainen *et al.*^{104,106}, with connections to community ecological theory and case studies. In our study, we modelled the virus community, denoted by the $n \times n_s$ matrix \mathbf{Y} of virus occurrences, comprising of individual components y_{ij} , denoting virus $j = 1, \dots, n_s$, where $n_s = 5$, on host plant $i = 1, \dots, n$, where $n = 320$, with probit regression

$$y_{ij} = 1_{L_{ij} + \varepsilon_{ij} > 0}$$

$$L_{ij} = L_{ij}^F + L_{ij}^R,$$

where $\varepsilon_{ij} \sim N(0,1)$, L_{ij} is the linear predictor for the occurrence of virus j on sentinel plant i , which is further divided to fixed (L_{ij}^F) and random (L_{ij}^R) parts. The fixed effects F model the influence of the local population context and the influence of the sentinel plant characteristics.

The random effect R models the residual variation in virus occurrences at the level of individual sentinel plants, that cannot be attributed to the above-described responses of the viruses to the fixed covariates. For exact formulation how the different components are modelled, with corresponding notation, please see Ovaskainen *et al.*¹⁰⁶.

Briefly, following the compact matrix notation of Chapter 7.3.2 in Ovaskainen *et al.*¹⁰⁶, we model the $n \times n_s$ community matrix of viruses \mathbf{Y} with a $n \times n_s$ matrix \mathbf{L} of all linear predictors L_{ij} for all species and sentinel plants, as $\mathbf{L} = \mathbf{L}^F + \mathbf{L}^R$. The matrix of fixed effects can be further decomposed as $\mathbf{L}^F = \mathbf{XB}$, where \mathbf{X} is the $n \times n_c$ matrix of environmental covariates, and \mathbf{B} is the $n_c \times n_s$ matrix of regression coefficients, i.e. species responses to the covariates, and n_c is the total amount of covariates included in the model. Because the environmental covariates \mathbf{X} are known and given as input for the model (Table 1), only the species responses \mathbf{B} are estimated. Analogously, the matrix of random effects can be decomposed as $\mathbf{L}^R = \mathbf{H}\mathbf{\Lambda}$. Here, \mathbf{H} is the $n \times n_f$ matrix of latent factors, or site loadings, and $\mathbf{\Lambda}$ is the $n_f \times n_s$ matrix of latent factor loadings, or, where n_f is the number of latent factors. Both the site \mathbf{H} and $\mathbf{\Lambda}$ are estimated, as is the number of latent factors n_f . The species loadings $\mathbf{\Lambda}$ can then be translated into residual associations between virus species by transforming them into covariation between species as $\mathbf{\Omega} = \mathbf{\Lambda}^T \mathbf{\Lambda}$, and further into correlations.

We fitted three JSJM variants to the data by varying the way the sentinel plant genotype was included in the model (Table 1). As explanatory variables (denoted by matrix \mathbf{X} in⁷¹) we used the local plant population context (categorical variable with four classes), which is a proxy for the plant population level effects such as variation in abiotic conditions, vector communities, and disease pool (categorical variable with four classes); and at the level of the sentinel host plants, we include the plant size (a continuous variable), signs of herbivory (a categorical variable with two classes; yes/no), as well the genotype of the sentinel host plant (a categorical variable with four classes). To examine the residual co-occurrence patterns among hosts, we also included the sentinel plant individual as a latent variable random effect.

First, we fitted a model with only the local population context, plant size and signs of herbivory (variant 1) as fixed explanatory variables \mathbf{X} . Then, we fitted a model including also the sentinel plant genotype, i.e. the full set of fixed explanatory variables (variant 2). With both of these model variants (1 and 2), we included random effects at the level of sentinel plants individuals. Finally, we fitted a model with the same full set of fixed explanatory variables \mathbf{X} as with model variant 2, but we modified the random effects by allowing these residual patterns to covary with the sentinel plant genotype (variant 3), details of which are explained by Tikhonov *et al.*¹⁰⁷. In this case, the latent factor loadings $\mathbf{\Lambda}$ are furthermore modelled as a linear regression of the selected fixed explanatory variables, which in this case was the sentinel plant genotype. Hence, as a summary, our model variants vary in terms of what is included in the matrix \mathbf{X} of explanatory variables, namely if sentinel plant genotype is included (variant 2) or not (variant 1), and do we allow the residual associations between viruses to covary among the sentinel genotype (variant 3) or not (variant 2).

We used the default priors of the package ‘Hmsc’¹⁰⁸, except that for the parameter $\mathbf{\Lambda}$ of species loadings, of the random part of the model. While the HMSC framework is usually not very sensitive to the choices of priors, when data is sufficient, they can be sensitive to the prior chosen for $\mathbf{\Lambda}$. The multiplicative gamma process shrinking prior¹⁰⁹ for the species loadings $\mathbf{\Lambda}$ has several prior parameters, but out of those, the user should pay attention to the choice of α , a vector of two values, which can be used to adjust the level of shrinkage that the prior implies for the matrix $\mathbf{\Omega}$ of species associations¹⁰⁶. Hence, we used two alternative priors. First, we used the default of $\alpha = (50,50)$, which imposes a lot of shrinkage. We refer to this group of model variants as our main model variants. Second, we used $\alpha = (3,3)$, which imposes much less shrinkage, but as a trade-off, also increases the risk of overfitting.

532

533 The model comparison approach allows us to examine the relevance of sentinel plant genotype
534 as a predictor of virus community composition (comparison of model variants 1 and 2), as well
535 as to see whether the residual co-occurrences between the viruses differ between the sentinel
536 plant genotypes (variant 3). The comparison of different priors enables us to examine how
537 sensitive our models were for these choices. We compared the model variants in terms of their
538 explanatory and predictive performance, where the first tells us how well the model predicts the
539 data used to fit it, whereas the latter illustrates how well the model predicts independent data
540 which has not been used for model fitting. We calculated the Tjur R^2 coefficient of
541 determination, a statistic that has been recommended to be used as a standard measure of
542 explanatory power for binary outcomes ¹¹⁰. The coefficient is obtained by calculating the mean
543 of the predicted probabilities of presences and absences, and then taking the difference between
544 those two means. Hence, a high coefficient value implies high predicted probabilities for
545 presences and low probabilities for absences. When interpreting it, it is good to note that with
546 sparse data, the probabilities of presence tend to be low in the first place, and thus the Tjur R^2
547 coefficient can remain rather low as well. Nevertheless, if the model is completely uninformative
548 and predicts a 50% probability for both presence and absence, the coefficient value will be zero,
549 thus revealing a poor model fit. For examining explanatory power, we fit the model to the full
550 data set and base our comparison on predictions made for the same data. To examine the
551 predictive power of the model, we conducted a 10-fold cross-validation and compared the model
552 variants based on the same Tjur R^2 coefficient as with explanatory power, but calculated from the
553 predictions made to new, unknown host plants. To complement our comparison based on model
554 accuracy, we calculated the widely applicable information criterion (WAIC)¹¹¹ for all the
555 variants.

556

We also conducted a partitioning of the variance explained by the best-performing model variant, to assess how different (groups of) variables are contributing to the overall variance explained by the model at the level of the linear predictor. Finally, we used the best-performing model variant to simulate predicted coinfections profiles.

We implemented our analyses with the R package ‘Hmsc’ (version 3.0-7¹⁰⁸). The performance comparison, variance partitioning and predictions were conducted with the tools provided in the package. For a full formal description of the structure of the modelling framework, please see Ovaskainen *et al.*^{104,106}, and for the covariate-dependent latent variables used in model variant 3, please see Tikhonov *et al.*¹⁰⁷. An R package including the analytical pipeline, data, and documentation for full reproduction of the results can be found in Github (aminorberg/trap17-pkg). For all the statistical analysis, we used R version 4.0.0¹¹². For more details on the statistical analysis, please see Supplementary Information (section ‘Supplementary information on the joint species distribution modelling’).

ACKNOWLEDGEMENTS

We thank Mikko Jalo, Pauliina Hyttinen and Vanja Milenkovic for helping with data collection in the field. We thank Pauliina Hyttinen and Laura Häkkinen for help with RNA-extractions. We thank Krista Raveala for help in cloning and caring for the plants. The Institute of Biotechnology at University of Helsinki is acknowledged for carrying out DNA-extractions. This research was

supported by funding from the Academy of Finland (296686) and European Research Council (4100097 RESISTANCE) to A-LL, and Luova Doctoral Programme Fellowship to SS.

CONTRIBUTIONS

SS, A-LL and HS designed the study. SS performed the experiment and data collection. AN performed the statistical analysis. SS, AN and A-LL wrote the manuscript and HS contributed substantially with comments on the manuscript.

DATA AVAILABILITY

The data supporting our results along with the analytical pipeline implemented as an R package are available in Github and accessible via <https://github.com/aminorberg/trap17-pkg>.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- 600 1. Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F. & Jetz, W. Homage to Linnaeus: How
601 many parasites? How many hosts? *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11482–11489 (2008).
- 602 2. Kreuze, J. F. *et al.* Complete viral genome sequence and discovery of novel viruses by deep
603 sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses.
604 *Virology* **388**, 1–7 (2009).
- 605 3. Prendeville, H. R., Ye, X., Jack Morris, T. & Pilson, D. Virus infections in wild plant populations
606 are both frequent and often unapparent. *Am. J. Bot.* **99**, 1033–1042 (2012).
- 607 4. Treena, I., B. *et al.* Distribution and diversity of Phytophthora across Australia Distribution and
608 diversity of Phytophthora across Australia. *Pacific Conserv. Biol.* **23**, 1–13 (2017).
- 609 5. Anderson, R. M. & May, R. M. Infectious Diseases of Humans; Dynamics and Control. (1991).
- 610 6. Hudson, P. J., Dobson, A. P. & Newborn, D. Prevention of population cycles by parasite removal.
611 *Science* **282**, 2256–2258 (1998).
- 612 7. Thrall, P. H. *et al.* Rapid genetic change underpins antagonistic coevolution in a natural host-
613 pathogen metapopulation. *Ecol. Lett.* **15**, 425–35 (2012).
- 614 8. Stevens, R. B. Cultural practices in disease control. in J. G. Horsfall and A. E. Dimond, editors.
615 Plant pathology, an advanced treatise. 357–429 (1960).
- 616 9. Susi, H., Barrès, B., Vale, P. F. & Laine, A.-L. Co-infection alters population dynamics of
617 infectious disease. *Nat. Commun.* **6**, 5975 (2015).
- 618 10. Connell, J. H. Diversity in tropical rain forests and coral reefs. *Science*. **199**, 1302–1310 (1978).
- 619 11. Tilman, D. Resource competition and community structure. 1–296 (1982).
- 620 12. Chesson, P. Mechanisms of Maintenance of Species Diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366
621 (2000).
- 622 13. Grime, P. J. Plant strategies, vegetation processes, and ecosystem properties. (2001).
- 623 14. Ovaskainen, O., Rybicki, J. & Abrego, N. What can observational data reveal about

metacommunity processes? *Ecography*. **42**, 1877–1886 (2019).

15. Seabloom, E. W., Hosseini, P. R., Power, A. G. & Borer, E. T. Diversity and composition of viral communities: coinfection of barley and cereal yellow dwarf viruses in California grasslands. *Am. Nat.* **173**, E79–E98 (2009).
16. Mihaljevic, J. R. Linking metacommunity theory and symbiont evolutionary ecology. *Trends Ecol. Evol.* **27**, 323–329 (2012).
17. Johnson, P. T. J., de Roode, J. C. & Fenton, A. Why infectious disease research needs community ecology. *Science* **349**, 1259504 (2015).
18. Borer, E. T., Laine, A.-L. & Seabloom, E. W. A Multiscale approach to plant disease using the metacommunity concept. *Annu. Rev. Phytopathol.* **54**, 397–418 (2016).
19. Tollenaere, C., Susi, H. & Laine, A.-L. Evolutionary and epidemiological implications of multiple infection in plants. *Trends Plant Sci.* **21**, 80–90 (2016).
20. Seabloom, E. W., Borer, E. T., Lacroix, C., Mitchell, C. E. & Power, A. G. Richness and composition of niche-assembled viral pathogen communities. *PLoS One* **8**, 1–9 (2013).
21. Borer, T., Seabloom, E. W., Mitchell, C. E. & Power, A. G. Local context drives infection of grasses by vector-borne generalist viruses. *Ecol. Lett.* **13**, 810–818 (2010).
22. Richgels, K. L. D., Hoverman, J. T. & Johnson, P. T. J. Evaluating the role of regional and local processes in structuring a larval trematode metacommunity of *Helisoma trivolvis*. *Ecography* **36**, 854–863 (2013).
23. Rodelo-Urrego, M. *et al.* Landscape heterogeneity shapes host-parasite interactions and results in apparent plant-virus codivergence. *Mol. Ecol.* **22**, 2325–2340 (2013).
24. Bernardo, P. *et al.* Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J.* **12**, 173–184 (2018).
25. Makiola, A. *et al.* Land use is a determinant of plant pathogen alpha- but not beta-diversity. *Mol.*

- 648 *Ecol.* **28**, 3786–3798 (2019).
- 649 26. Cottenie, K. Integrating environmental and spatial processes in ecological community dynamics.
650 *Ecol. Lett.* **8**, 1175–1182 (2005).
- 651 27. Dobson, A. Population dynamics of pathogens with multiple host species. *Am. Nat.* **164**, 64–78
652 (2004).
- 653 28. Malpica, M. & Sacrista, S. Association and host selectivity in multi-host pathogens. *PLoS One* **1**,
654 e41 (2006).
- 655 29. Poulin, R., Blanan, C. A., Thieltges, D. W. & Marcogliese, D. J. Scaling up from epidemiology to
656 biogeography: Local infection patterns predict geographical distribution in fish parasites. *J.*
657 *Biogeogr.* **39**, 1157–1166 (2012).
- 658 30. Nunn, C. L., Altizer, S., Jones, K. E. & Sechrest, W. Comparative tests of parasite species richness
659 in primates. *Am. Nat.* **162**, 687–614 (2003).
- 660 31. Ezenwa, V. O., Price, S. A., Altizer, S., Vitone, N. D. & Cook, K. C. Host traits and parasite
661 species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. *Oikos*
662 **115**, (2006).
- 663 32. Gilbert, G. S. & Webb, C. O. Phylogenetic signal in plant pathogen – host range. *Proc. Natl. Acad.*
664 *Sci. U. S. A.* **104**, 4979–4983 (2007).
- 665 33. Lindenfors, P. *et al.* Parasite species richness in carnivores: effects of host body mass, latitude,
666 geographical range and population density. *Glob. Ecol. Biogeogr.* **16**, 496–509 (2007).
- 667 34. Mitchell, C. E., Blumenthal, D., Jarošík, V., Puckett, E. E. & Pyšek, P. Controls on pathogen
668 species richness in plants’ introduced and native ranges: Roles of residence time, range size and
669 host traits. *Ecol. Lett.* **13**, 1525–1535 (2010).
- 670 35. Endara, M.-J. & Coley, P. D. The resource availability hypothesis revisited: a meta-analysis.
671 *Funct. Ecol.* **25**, 389–398 (2011).

- 672 36. Miller, Z. J. Notes and Comments Fungal Pathogen Species Richness: Why do some plant species
673 have more pathogens than others? *Am. Nat.* **179**, 282–292 (2012).
- 674 37. Dallas, T. A. & Presley, S. J. Relative importance of host environment, transmission potential and
675 host phylogeny to the structure of parasite metacommunities. *Oikos* **123**, 866–874 (2014).
- 676 38. Strong, D. R. & Levin, D. A. Species Richness of Plant Parasites and Growth Form of Their Hosts
677 Author (s): Donald R . Strong , Jr . and Donald A . Levin Published by : The University of
678 Chicago Press for The American Society of Naturalists Stable URL : <https://www.jstor.org/stable>.
679 *Am. Nat. Nat.* **114**, 1–22 (2019).
- 680 39. Louhi, K. R., Karvonen, A., Rellstab, C., Louhi, R. & Jokela, J. Prevalence of infection as a
681 predictor of multiple genotype infection frequency in parasites with multiple-host life cycle. *J.*
682 *Anim. Ecol.* **82**, 191–200 (2013).
- 683 40. Raybould, A. F., Maskell, L. C., Edwards, M. L., Cooper, J. I. & Gray, A. J. The prevalence and
684 spatial distribution of viruses in natural populations of *Brassica oleracea*. *New Phytol.* **141**, 265–
685 275 (1999).
- 686 41. Shoemaker, L. G. *et al.* Pathogens manipulate the preference of vectors, slowing disease spread in
687 a multi-host system. *Ecol. Lett.* **22**, 1115–1125 (2019).
- 688 42. Shaw, D. J. & Dobson, a P. Patterns of macroparasite abundance and aggregation in wildlife
689 populations: a quantitative review. *Parasitology* **111 Suppl**, S111–S127 (1995).
- 690 43. Klein, J., Satta, Y. & Uigin, O. The molecular descent of the major histocompatibility complex.
691 *Annu. Rev. Immunol.* **11**, 269–95 (1993).
- 692 44. Kamitani, M., Nagano, A. J., Honjo, M. N. & Kudoh, H. RNA-Seq reveals virus-virus and virus-
693 plant interactions in nature. *FEMS Microbiol. Ecol.* **92**, 1–11 (2016).
- 694 45. Mysore, K. S. & Ryu, C. M. Nonhost resistance: How much do we know? *Trends Plant Sci.* **9**, 97–
695 104 (2004).
- 696 46. Remold, S. K. Unapparent virus infection and host fitness in three weedy grass species. *J. Ecol.*

- 697 967–977 (2002).
- 698 47. Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. Evolutionary Dynamics of Plant R -Genes.
699 *Plant Pathol.* **292**, 2281–2286 (2001).
- 700 48. Laine, A.-L. Detecting local adaptation in a natural plant-pathogen metapopulation: a laboratory
701 vs. field transplant approach. *J. Evol. Biol.* **20**, 1665–73 (2007).
- 702 49. Mandadi, K. K. & Scholthof, K. B. G. Plant immune responses against viruses: How does a virus
703 cause disease? *Plant Cell* **25**, 1489–1505 (2013).
- 704 50. Bruns, E., Carson, M. & May, G. Pathogen and host genotype differently affect pathogen fitness
705 through their effects on different life-history stages. *BMC Evol. Biol.* **12**, 135 (2012).
- 706 51. Strauss, A. T., Bowling, A. M., Duffy, M. A., Cáceres, C. E. & Hall, S. R. Linking host traits,
707 interactions with competitors and disease: Mechanistic foundations for disease dilution. *Funct.*
708 *Ecol.* **32**, 1271–1279 (2018).
- 709 52. Mundt, C. C. Use of Multiline Cultivars and Cultivar Mixtures for Disease Management. *Annu.*
710 *Rev. Phytopathol.* **40**, 381–410 (2002).
- 711 53. Jousimo, J. *et al.* Disease ecology. Ecological and evolutionary effects of fragmentation on
712 infectious disease dynamics. *Science* **344**, 1289–93 (2014).
- 713 54. Ekroth, A. K. E., Rafaluk-Mohr, C. & King, K. C. Host genetic diversity limits parasite success
714 beyond agricultural systems: a meta-analysis. *Proc. R. Soc. B Biol. Sci.* **286**, 20191811 (2019).
- 715 55. Bolnick, D. I. *et al.* Why intraspecific trait variation matters in community ecology. *Trends Ecol.*
716 *Evol.* **26**, 183–192 (2011).
- 717 56. Rausher, M. D. Tradeoffs in Performance on Different Hosts: Evidence from Within- and
718 Between-Site Variation in the Beetle *Deloyala guttata*. *Evolution* **38**, 582–595 (1984).
- 719 57. Stearns, S. C. Trade-Offs in Life-History Evolution. *Funct. Ecol.* **3**, 259–268 (1989).
- 720 58. Zhou, X. *et al.* Loss of function of a rice TPR-domain RNA-binding protein confers broad-

721 spectrum disease resistance. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3174–3179 (2018).

722 59. Karvonen, A., Jokela, J. & Laine, A. L. Importance of Sequence and Timing in Parasite
723 Coinfections. *Trends Parasitol.* **35**, 109–118 (2019).

724 60. Van Hulten, M., Pelser, M., Van Loon, L. C., Pieterse, C. M. J. & Ton, J. Costs and benefits of
725 priming for defense in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5602–5607 (2006).

726 61. Graham, A. L. Ecological rules governing helminth-microparasite coinfection. *Proc. Natl. Acad.*
727 *Sci. U. S. A.* **105**, 566–570 (2008).

728 62. Laine, A.-L. Context-dependent effects of induced resistance under co-infection in a plant-
729 pathogen interaction. *Evol. Appl.* **4**, 696–707 (2011).

730 63. Syller, J. Facilitative and antagonistic interactions between plant viruses in mixed infections. *Mol.*
731 *Plant Pathol.* **13**, 204–216 (2012).

732 64. Porrozzzi, R., Teva, A., Amaral, V. F., Santos Da Costa, M. V. & Grimaldi, G. Cross-immunity
733 experiments between different species or strains of Leishmania in rhesus macaques (Macaca
734 mulatta). *Am. J. Trop. Med. Hyg.* **71**, 297–305 (2004).

735 65. Burdon, A. J. J., Thrall, P. H. & Brown, A. H. D. Resistance and Virulence Structure in Two
736 *Linum marginale*-*Melampsora lini* Host-Pathogen Metapopulations with Different Mating
737 Systems. *Evolution* **53**, 704–716 (1999).

738 66. Lively, C. M., de Roode, J. C., Duffy, M. A., Graham, A. L. & Koskella, B. Interesting open
739 questions in disease ecology and evolution. *Am. Nat.* **184**, (2014).

740 67. Roossinck, M. J. & Bazán, E. R. Symbiosis: Viruses as Intimate Partners. *Annu. Rev. Virol.* **4**,
741 123–139 (2017).

742 68. Hily, J. M., Poulicard, N., Mora, M. Á., Pagán, I. & García-Arenal, F. Environment and host
743 genotype determine the outcome of a plant-virus interaction: From antagonism to mutualism. *New*
744 *Phytol.* **209**, 812–822 (2016).

- 745 69. Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P. & Laine, A.-L. Diverse and variable virus
746 communities in wild plant populations revealed by metagenomic tools. *PeerJ* **2019**, e6140 (2019).
- 747 70. Warton, D. I. *et al.* So Many Variables: Joint Modeling in Community Ecology. *Trends Ecol.*
748 *Evol.* **30**, 766–779 (2015).
- 749 71. Ovaskainen, O. & Soininen, J. Making more out of sparse data: hierarchical modeling of species
750 communities. *Ecology* **92**, 289–295 (2011).
- 751 72. Norberg, A. *et al.* A comprehensive evaluation of predictive performance of 33 species
752 distribution models at species and community levels. *Ecol. Monogr.* **89**, e01370 (2019).
- 753 73. Aivelo, T. & Norberg, A. Parasite–microbiota interactions potentially affect intestinal
754 communities in wild mammals. *J. Anim. Ecol.* **87**, (2018).
- 755 74. Dallas, T. A., Laine, A., Ovaskainen, O. & Dallas, T. A. Detecting parasite associations within
756 multi-species host and parasite communities. *Proc. R. Soc. B Biol. Sci.* **286**, 20191109 (2019).
- 757 75. Alizon, S., de Roode, J. C. & Michalakis, Y. Multiple infections and the evolution of virulence.
758 *Ecol. Lett.* **16**, 556–567 (2013).
- 759 76. Thrall, P. H. & Burdon, J. J. Effect of resistance variation in a natural plant host-pathogen
760 metapopulation on disease dynamics. *Plant Pathol.* **49**, 767–773 (2000).
- 761 77. Thrall, P. H., Burdon, J. J. & Young, A. Variation in resistance and virulence among demes of a
762 plant host-pathogen metapopulation. *J. Ecol.* **89**, 736–748 (2001).
- 763 78. Lively, C. M. The effect of host genetic diversity on disease spread. *Am. Nat.* **175**, (2010).
- 764 79. Benavides, J. A. *et al.* From Parasite Encounter to Infection: Multiple-scale drivers of parasite
765 richness in a wild social primate population. *Am. J. Phys. Anthropol.* **147**, 52–63 (2012).
- 766 80. Susi, H. & Laine, A. L. The effectiveness and costs of pathogen resistance strategies in a perennial
767 plant. *J. Ecol.* **103**, 303–315 (2015).
- 768 81. Ojanen, S. P., Nieminen, M., Meyke, E., Pöyry, J. & Hanski, I. Long-term metapopulation study

- 769 of the Glanville fritillary butterfly (*Melitaea cinxia*): Survey methods, data management, and long-
770 term population trends. *Ecol. Evol.* **3**, 3713–3737 (2013).
- 771 82. Jackson, J. A., Pleass, R. J., J, C., Bradley, J. E. & Tinsley, R. C. Heterogenous interspecific
772 interactions in a host – parasite system. *Int. J. Parasitol.* **36**, 1341–1349 (2006).
- 773 83. Susi, H., Vale, P. F. & Laine, A.-L. Host Genotype and coinfection modify the relationship of
774 within and between host transmission. *Am. Nat.* **186**, 000–000 (2015).
- 775 84. Laine, A. Resistance variation within and among host populations in a plant – pathogen
776 metapopulation : implications for. *J. Ecol.* **92**, 990–1000 (2004).
- 777 85. Greischar, M. A. & Koskella, B. A synthesis of experimental work on parasite local adaptation.
778 *Ecol. Lett.* **10**, 418–434 (2007).
- 779 86. Hoeksema, J. D. & Forde, S. E. A meta-analysis of factors affecting local adaptation between
780 interacting species. *Am. Nat.* **171**, 275–290 (2008).
- 781 87. Laine, A.-L., Burdon, J. J., Dodds, P. N. & Thrall, P. H. Spatial variation in disease resistance:
782 from molecules to metapopulations. *J. Ecol.* **99**, 96–112 (2011).
- 783 88. Smith, C. M. & Boyko, E. V. The molecular bases of plant resistance and defense responses to
784 aphid feeding : current status. *Entomol. Exp. Appl.* **122**, 1–16 (2007).
- 785 89. Crutsinger, G. M. *et al.* Plant genotypic diversity predicts community structure and governs an
786 Ecosystem Process. *Science (80-.)*. **647**, 966–968 (2006).
- 787 90. Mauck, K., Bosque-Pérez, N. A., Eigenbrode, S. D., De Moraes, C. M. & Mescher, M. C.
788 Transmission mechanisms shape pathogen effects on host-vector interactions: Evidence from plant
789 viruses. *Funct. Ecol.* **26**, 1162–1175 (2012).
- 790 91. Susi, H. & Laine, A. L. Host resistance and pathogen aggressiveness are key determinants of
791 coinfection in the wild. *Evolution (N. Y.)*. **71**, 2110–2119 (2017).
- 792 92. Bergner, L. M. *et al.* Demographic and environmental drivers of metagenomic viral diversity in

793 vampire bats. *Mol. Ecol.* **29**, 26–39 (2019).

794 93. Telfer, S. *et al.* Species Interactions in a Parasite Community Drive Infection Risk in a Wildlife
795 Population. *Science*. **330**, 243–247 (2010).

796 94. Halliday, F. W. *et al.* Facilitative priority effects drive parasite assembly under coinfection. *Nat.*
797 *Ecol. Evol.* [doi.org/10](https://doi.org/10.1038/s41559-020-1100-0), (2020).

798 95. Mauch-Mani, B., Baccelli, I., Luna, E. & Flors, V. Defense priming: an adaptive part of induced
799 resistance. *Annu. Rev. Plant Biol.* **68**, 485–512 (2017).

800 96. Cassells, A. C. & Herrick, C. C. Cross protection between mild and severe strains of tobacco
801 mosaic virus in doubly inoculated tomato plants. *Virology* **78**, 253–260 (1977).

802 97. Sagar, A. G. R. & Harper, J. L. *Plantago Major L.*, *P. Media L.* and *P. Lanceolata L.* *J. Ecol.* **52**,
803 189–221 (1964).

804 98. Laine, A.-L. & Hanski, I. Large-scale spatial dynamics of a specialist plant pathogen in a
805 fragmented landscape. *J. Ecol.* **94**, 217–226 (2006).

806 99. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356–363 (2004).

807 100. Chang, S., Puryear, J. & Cairny, J. A simple and efficient method for isolating RNA from pine
808 trees. *Plant Mol. Biol. Report.* **11**, 113–116 (1993).

809 101. Susi, H. *et al.* Genome sequences of a capulavirus infecting *Plantago lanceolata* in the Åland
810 archipelago of Finland. *Arch. Virol.* **162**, 2041–2045 (2017).

811 102. Griffith, D. M., Veech, J. A. & Marsh, C. J. cooccur : Probabilistic Species Co-Occurrence
812 Analysis in R. *J. Stat. Softw.* **69**, 1–17 (2016).

813 103. Veech, J. A. A probabilistic model for analysing species co-occurrence. *Glob. Ecol. Biogeogr.* **22**,
814 252–260 (2013).

815 104. Ovaskainen, O. *et al.* How to make more out of community data? A conceptual framework and its
816 implementation as models and software. *Ecol. Lett.* **2**, 561–576 (2017).

- 817 105. Elith, J. & Leathwick, J. R. Species Distribution Models: Ecological Explanation and Prediction
818 Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
- 819 106. Ovaskainen, O. & Abrego, N. *Joint Species Distribution Modelling: With Applications in R.*
820 *Ecology, Biodiversity and Conservation* (Cambridge University Press, 2020).
- 821 107. Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. Using joint species distribution models
822 for evaluating how species-to-species associations depend on the environmental context. *Methods*
823 *Ecol. Evol.* **8**, 443–452 (2017).
- 824 108. Tikhonov, G. *et al.* Hmsc: Hierarchical Model of Species Communities. R package version 3.0-7.
825 (2020).
- 826 109. Bhattacharya, A. & Dunson, D. B. Sparse Bayesian infinite factor models. *Biometrika* **98**, 291–
827 306 (2011).
- 828 110. Tjur, T. Coefficients of Determination in Logistic Regression Models—A New Proposal: The
829 Coefficient of Discrimination. *Am. Stat.* **63**, 366–372 (2009).
- 830 111. Watanabe, S. A widely applicable bayesian information criterion. *J. Mach. Learn. Res.* **14**, 867–
831 897 (2013).
- 832 112. R Core Team. R: A Language and Environment for Statistical Com- puting. R Foundation for
833 Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. (2020).

834

835

Table 1. Joint species distribution model variants and their explanatory performance and predictive performance (based on cross-validation), measured by the Tjur R^2 coefficient of determination¹¹⁰ (see Methods and Supplementary Information).

Model variant	Fixed explanatory variables	Random effects (latent variables)	Explanatory performance	Predictive performance	WAIC
1	Local population context, host size, signs of herbivory	Host plant individual	0.072	0.041	2.00
2	Host genotype, local population context, host size, signs of herbivory	Host plant individual	0.16	0.11	1.78
3	Host genotype, local population context, host size, signs of herbivory	Genotype-dependent host plant individual	0.16	0.11	1.81

Table 2. Regression coefficients for model variant 2 for each virus species. Posterior mean estimates with statistical support based on the 90% central credible interval are denoted by bold font. ‘Clo’ refers to *Plantago closterovirus*, ‘Be’ to *Plantago betapartitivirus*, ‘Cap’ to *Plantago lanceolate latent virus*, ‘Cau’ to *Plantago latent caulimovirus*, and ‘En’ refers to *Plantago enamovirus*.

	Clo	Be	Cap	Cau	En
(Intercept)	-11	-1.2	-1.8	-1.9	-2.0
Host plant size	0.00094	0.00042	0.00056	0.00017	-0.00038
Signs of herbivory	0.48	0.30	-0.22	-0.29	-0.21
Population 9031	0.30	0.13	0.053	-0.47	0.27
Population 3302	-0.20	0.57	0.62	0.18	-0.084
Population 433	-0.0070	-0.091	-0.14	-0.26	-0.58
Genotype 609_19	0.12	0.12	2.47	0.67	0.067
Genotype 4_13	0.017	0.23	0.46	-0.011	0.35
Genotype 2818_6	-0.16	-0.081	0.23	-0.25	-0.57

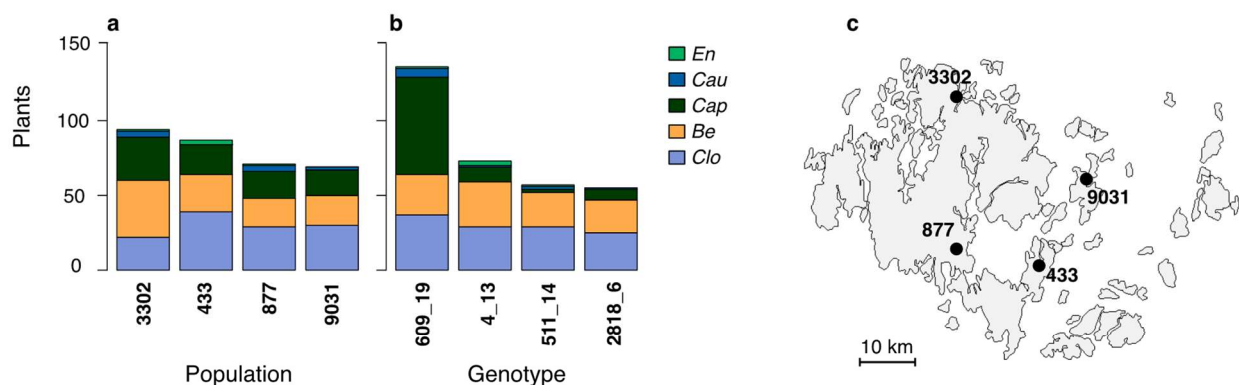
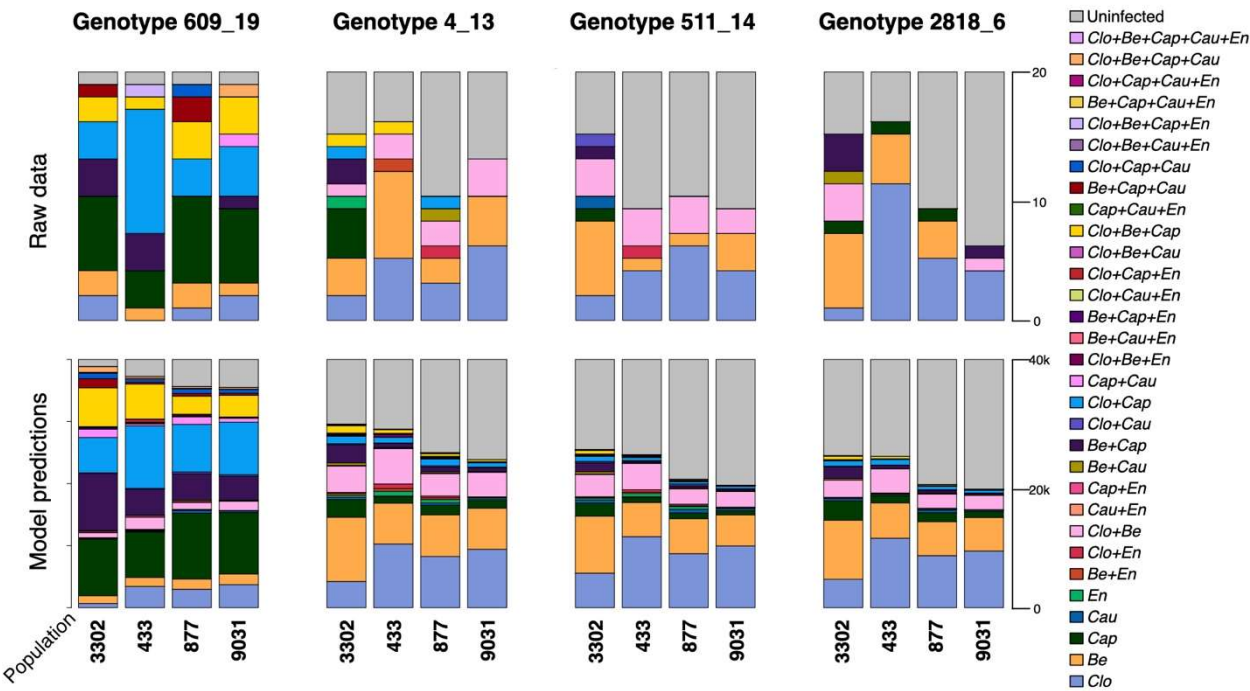


Figure 1. Virus infections in sentinel plants. Infections are plotted by population (a) and genotype (b), and the locations of the study populations in the field experiment in the Åland Islands (c). The genotypes and populations are ordered from left to right according to decreasing overall number of infections. ‘Clo’ refers to *Plantago closterovirus*, ‘Be’ to *Plantago betapartitivirus*, ‘Cap’ to *Plantago lanceolate latent virus*, ‘Cau’ to *Plantago latent caulimovirus*, and ‘En’ refers to *Plantago enamovirus*.

855



856

857

858

859

860

861

862

863

864

865

866

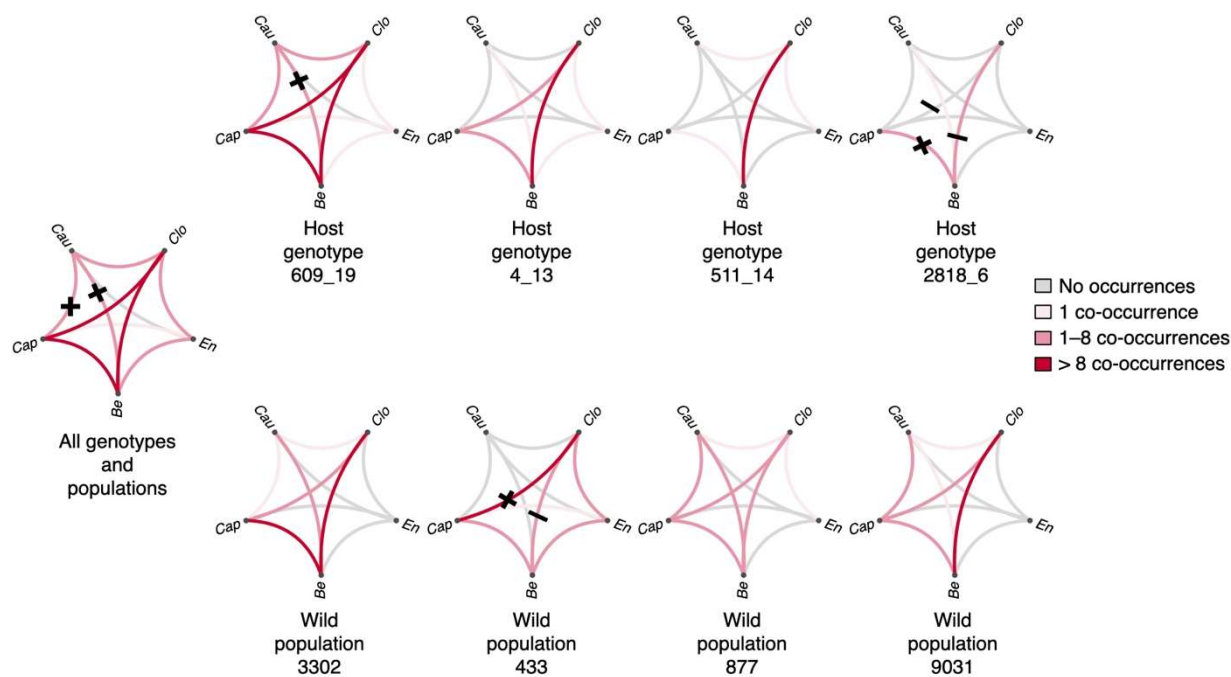
867

868

869

Figure 2. (Co-)infections in the original data (upper panel) and predicted coinfections (lower panel) based on the model variant 2, ordered as host genotypes for each population, as indicated by the horizontal axis. Both the genotypes and populations are ordered with respect to frequency, so that the bars on the left-hand side show the population and genotype with the highest total amount of virus infection. ‘Clo’ refers to *Plantago closterovirus*, ‘Be’ to *Plantago betapartitivirus*, ‘Cap’ to *Plantago lanceolate latent virus*, ‘Cau’ to *Plantago latent caulimovirus*, and ‘En’ refers to *Plantago enamovirus*. The total number of plants in the upper panel is 20, whereas in the lower panel the total number is simulated plants is 20 (original number of plants) times 2000 (number of MCMC iterations used for the simulation), resulting in 40000 simulated plants.

870



871

872 **Figure 3. Co-occurrences between virus species.** Co-occurrences are shown either in the whole
873 data set (left, with total number of sentinel plants 320), or per plant genotype (upper panels, 80
874 plants per genotype), or by population (lower panels, 80 plants per population) as denoted by the
875 horizontal axis. The genotypes and populations are ordered from left to right according to
876 decreasing overall frequency of disease. The plus (and minus) signs denote the pairs, for which
877 the observed values were higher (or lower, respectively) than what would be expected based on
878 their overall frequencies, and for which the probability of this difference was < 0.1 . The line
879 colours denote the true numbers of co-occurrences between the species, as shown in the legend.
880 'Clo' refers to *Plantago closterovirus*, 'Be' to *Plantago betapartitivirus*, 'Cap' to *Plantago*
881 *lanceolate latent virus*, 'Cau' to *Plantago latent caulimovirus*, and 'En' refers to *Plantago*
882 *enamovirus*. The exact probabilities for the focal pairs are provided in the Supplementary table 3.

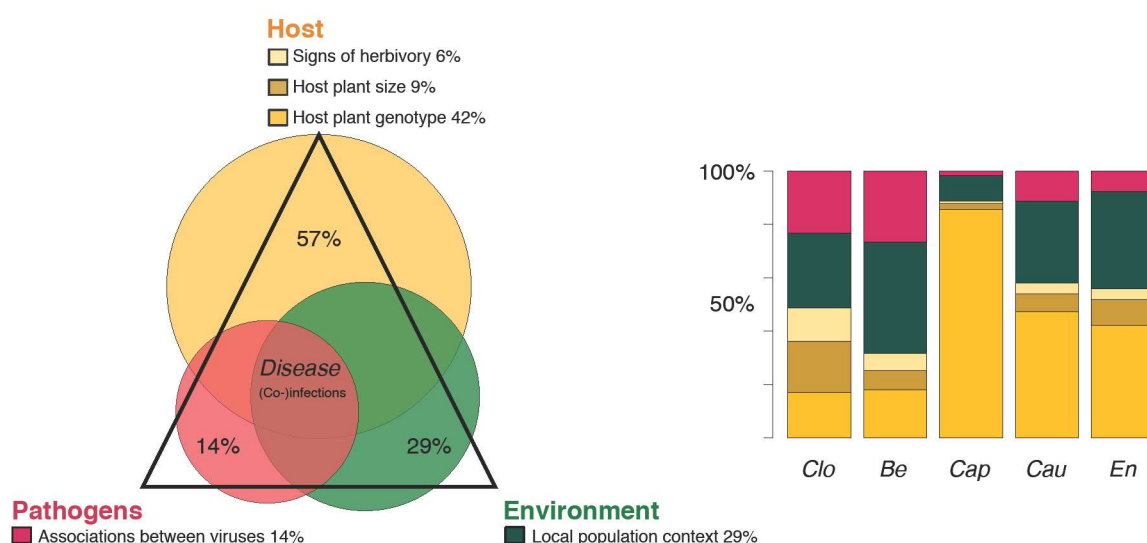


Figure 4. Partitioning of the variance explained by model variant 2 (Table 1). The diagram overlays the average proportions (over species) of variance explained by different groups of explanatory variables (out of the total variation explained by the model) and the concept of the disease triangle. The legend labels denote the different variables for which the partitioning is calculated, and the percentages indicate the mean values for the whole community. The barplot gives these results separately for each virus: the horizontal axis shows the focal five viruses (ordered from left to right according to their decreasing overall infection rate) and the vertical axis shows the proportion of variance explained. ‘Clo’ refers to *Plantago closterovirus*, ‘Be’ to *Plantago betapartitivirus*, ‘Cap’ to *Plantago lanceolate latent virus*, ‘Cau’ to *Plantago latent caulimovirus*, and ‘En’ refers to *Plantago enamovirus*.